

## **TITLE PAGE**

**Title:** PHONE AND VIDEO-BASED MODALITIES OF CENTRAL BLINDED ADJUDICATION OF MODIFIED RANKIN SCORES IN AN ENDOVASCULAR STROKE TRIAL

**Authors:** Elena López-Cancio (MD, PhD), Mercè Salvat (N), Neus Cerdà (MS), Marta Jiménez (MD), Javier Codas (MD), Laura Llull (MD), Sandra Boned (MD), Luis M. Cano (MD), Blanca Lara (MD), Carlos Molina (MD, PhD), Erik Cobo (MD, PhD), Antoni Dávalos (MD, PhD), Tudor G. Jovin (MD), Joaquín Serena (MD, PhD), On Behalf of REVASCAT investigators

**Author's affiliations:** From the Department of Neuroscience, Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Barcelona, Spain (E.L-C, M.J, A.D); Stroke Program, Agency for Health Quality and Assessment of Catalonia, Spain (M.S); Bioclever, Department of Statistics, Barcelona, Spain (N.C); Hospital Clínic i Provincial, Barcelona, Spain (J.C, L.L); Hospital Vall d'Hebron, Barcelona, Spain (S.B, C.M); Hospital de Bellvitge, L'Hospitalet de Llobregat, Barcelona, Spain (B.L, L.M.C); Department of Biostatistics, Universitat Politècnica de Catalunya, Barcelona, Spain (E.C); Stroke Institute, Department of Neurology, UPMC, Pittsburgh, Pennsylvania, USA (T.G.J.); Stroke Unit, Hospital Josep Trueta, Girona, Spain (J.S)

**Corresponding Author:** Elena López-Cancio (MD, PhD)

Adress: Department of Neuroscience, Stroke Unit, Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona. Ctra Canyet s/n, 08916 Badalona, Barcelona, Spain

Phone: 34.934978911

Fax: 34.934978742

e-mail: [elenacancio@gmail.com](mailto:elenacancio@gmail.com)

**Cover title:** Blinded central adjudication of mRS in REVASCAT

**Tables and figures:** 1 figure and 4 tables

**Key words:** endovascular, stroke trial, Rankin, central adjudication

**Subject codes:** 44, 83

**Total word count** (including title page, abstract, main body of text, acknowledgments, sources of funding, disclosures, references, figure legends, tables): 3971

# PHONE AND VIDEO-BASED MODALITIES OF CENTRAL BLINDED ADJUDICATION OF MODIFIED RANKIN SCORES IN AN ENDOVASCULAR STROKE TRIAL

## Abstract

**Background and purpose:** The standard outcome measure in stroke research is modified Rankin Scale (mRS) evaluated by local blinded investigators. We aimed to assess feasibility and reliability of two central adjudication methods of mRS in the setting of a randomized endovascular stroke trial.

**Methods:** This is a secondary analysis derived from the REVASCAT trial cohort. Primary outcome was distribution of mRS at 90 days. Local evaluation was done by certified investigators masked to treatment assignment using structured face-to-face interviews. In addition, central assessment was performed by two independent raters via structured phone interview (n=120) and via video-recordings of the face-to-face interviews with local investigators (n=106). Interrater agreement was evaluated using kappa and discordance statistics. Sensitivity analyses for the primary endpoint using different adjudication approaches were performed. Correlation between mRS obtained with each modality and 24 hour follow-up infarct volumes (FIV) was studied.

**Results:** Using local evaluation as the reference, higher agreement rates were noted with central video than with central phone evaluations ( $k_w$  0.92[0.88-0.96] versus 0.77[0.72-0.83]). Discrepancies in mRS scoring between local and central raters (phone and video-based) were similar in both treatment allocation arms. Sensitivity analyses showed benefit of endovascular treatment irrespective of adjudication method, but higher odds ratios were observed with local evaluations. FIV was similarly correlated with mRS across all three evaluation modalities.

**Conclusion:** Central adjudication of mRS is feasible, reducing interrater variability and avoiding potential problems related to lack of blinding. Our findings may have implications in the planning of future randomized acute stroke trials, especially in those including non-pharmacological interventions.

**Clinical trial registration:** URL: <http://www.clinicaltrials.gov>. Unique identifier: NCT01692379

## Introduction

The modified Rankin scale (mRS), the standard outcome measure in acute stroke research, is usually assessed by local blinded investigators through in person encounters<sup>1</sup>. Although core labs for adjudication of secondary end points (eg. imaging) are frequently used in stroke research, central adjudication of mRS scores is less commonly used. There is a substantial interobserver variability in mRS assessment that persists even with certified assessors and using structured interviews<sup>2,3,4</sup>. For open label trials such as procedure based stroke studies, an additional shortcoming is the difficulty of local investigators to remain blinded to treatment allocation.

Central or external evaluation of functional outcome may overcome these limitations. Telephonic interviews assessing mRS have shown a modest agreement with face-to-face interviews in exploratory studies<sup>5, 6</sup>, and this agreement has not been assessed properly in a clinical trial setting. Recently, Mc Arthur and colleagues published the feasibility and reliability of a video-based modality for central remote mRS adjudication in a virtual multicenter stroke trial using a group adjudication approach with four central assessors<sup>7</sup>. Agreement between the centrally adjudicated and local evaluations was good. They also demonstrated the feasibility of using translated interviews simulating an international multicenter study. However, the use of external video-based adjudication methodology in a real-world trial has not been published yet.

In order to address the above mentioned concerns related to in person evaluation, REVASCAT (Randomized Trial of Revascularization with Solitaire FR® device versus Best Medical Therapy in the Treatment of Acute Stroke due to Anterior Circulation Large Vessel Occlusion Presenting within Eight Hours of Symptom Onset) was designed to include both local and central evaluation methods of the primary outcome. This secondary analysis from REVASCAT study aimed to determine feasibility of central mRS adjudication and compare different central adjudication methods (phone and video-based). In addition, since the true disability status cannot be determined, the three methods of

outcome adjudication were also compared with a more objective measure of cerebral impairment, core lab evaluated 24 hour infarct volume on CT or MRI scans.

## **Methods**

### **Study design**

REVASCAT enrolled 206 patients, randomized to thrombectomy with Solitaire device versus medical management alone. Eligible patients had contraindications to IV alteplase or had received IV alteplase therapy within 4.5 hours without revascularization after 30 minutes of alteplase infusion. Primary endpoint was distribution of mRS scores at 90 days ( $\pm 14$  days). Details on study protocol and main results of the study have been already published<sup>8,9</sup>.

### **Modified Rankin scale evaluations**

The primary outcome variable was evaluated twice in each patient by both local and central certified assessors. Locally, each site designated one or more mRS-certified neurologists, not involved in patient management, to evaluate the mRS score in a face-to-face visit. Local investigators were asked to follow a specific structured interview based on the Rankin Focused Assessment<sup>10</sup>.

Methodology for central adjudication of mRS scores varied along REVASCAT study enrolment: during the first period of the study an external mRS-certified nurse (MS) evaluated mRS scores by telephone call using the same structured interview as local investigators and recorded the interviews by audio-tape. After 50 patients centrally adjudicated in this manner, based on a pre-determined unblinded safety analysis, the DSMB recommended implementing measures that would result in a higher concordance between local and centrally adjudicated outcomes. For that reason the steering committee decided to implement a remote video-based adjudication of mRS. After the approval of the ethical committee of each center, the local face-to-face structured interviews were video recorded and stored in a specific laptop equipped with a portable web-cam and internet connection

that was distributed to each center with this purpose. Video-clips were uploaded and transferred via file transfer protocol (FTP) to one single external, mRS-certified neurologist (JS). For the ensuing 55 patients and in order to ensure continuity of central assessments in case the video evaluation proved to be unfeasible, both central video and central audio evaluations were performed. Central assessors were located outside the enrolling hospitals, were not aware of mRS scores given by local investigators, and introduced their assessments in an independent part of eCRF.

Central video rater also evaluated, in a blinded-manner, quality of face-to-face interviews performed by local investigators and scored them as: 1. Poor (local investigator did not follow structured interview systematically, central rater had to display video-clip more than two times to decide on a final assessment); 2. Acceptable (structured interview is used but central rater has some uncertainties in evaluation between two adjacent items, having to view part of the video-clip at least twice); 3. Reliable (structured interview is completely followed and central rater has no hesitation in ascribing a final adjudication after reviewing video-clip only once).

### **Final infarct volume measurement**

Infarct volumes at 24 hours on CT or MRI were adjudicated blinded to clinical data by image core lab. In case both CT and MRI were available, for uniformity sake CT data was used. Volumes were calculated through software generated volumetric calculations (Quantomo, <sup>11</sup>) with manual adjustments when deemed necessary. In case of coexistence of hemorrhage and infarction, combined (infarction and hemorrhage) volumes were used.

### **Statistical analysis**

Global inter-rater agreement in mRS scoring between local and central raters was evaluated with weighted kappa statistics. Magnitude and direction of discrepancies (difference between central rater and local rater) were also studied. In order to detect possible bias regarding lack of blinding, those discrepancies were compared between both treatment allocation arms with T test, as sample size was big enough to employ convergence to Normal distribution. First, discrepancies were evaluated globally between local and central raters, including phone and video-based raters (in the 55 patients with both central assessments, the average score [phone+video]/2 was used). Thereafter, due to the different

methodology between phone and video-based evaluations, discrepancies of local investigators with each of central assessors was evaluated also separately. All analyses were performed excluding deaths (mRS=6). In patients evaluated by the central video-based rater, percentage of total agreement with local evaluations was also studied taking into account quality of interviews (poor, acceptable, reliable).

Primary analysis in REVASCAT was based on central evaluator through video recording with local investigator's assessment as the default modality in case the video evaluation was not available as per steering committee decision prior to data unblinding. Sensitivity analyses for the main outcome were also performed based on scores given by both central evaluators (sensitivity analysis I) and local investigators (sensitivity analysis II). All analyses were in the intention-to-treat population. Effect size measure was a cumulative logistic regression odds ratio (shift analysis) and 95% confident interval, adjusted for minimization factors and IV alteplase use.

Correlation (Spearman) of final infarct volume with mRS scores obtained with each of the three mRS assessments was performed.

## **Results**

From November 2012 to December 2014, 206 patients were enrolled in REVASCAT trial in four centers in Catalonia, Spain. All patients had available outcomes at 90 days performed via at least two methods. Excluding deaths, 171 patients had an evaluation of mRS at 90 days performed by masked local investigators at each site (in 5 patients the local evaluation was made by phone due to patient's impossibility to attend the hospital). During the first period of the study, 120 patients also underwent a central phone based assessment of functional status at 90 days. In the second part of the study, 106 patients were video-recorded during the face-to-face interviews at local sites, and a single central assessor evaluated mRS in video-recordings. There were 55 patients that received all three evaluations (local, phone-based and video-based).

### **Feasibility of central adjudication of mRS during the trial**

All patients or proxies consented to mRS evaluation by any method, anonymity was maintained in all patients and the files containing the functional assessment (audio and

video clips) were stored as back-up copies. For central phone adjudication, a notification was sent via fax to the central evaluator including the name of the patient and a direct relative, contact phones and the date to be called after randomization. Phone calls were made from a central office of the Catalan Stroke Program. Remote video-based assessment required the implementation of specific technology in the four participating centers consisting on a light laptop equipped with web-cam and internet connectivity to directly store the video-clips and afterwards upload them to the transmission system. The portability of the laptop facilitated that the interviews could be made in any outpatient office. Video protocol was brief and no technical issues were encountered with respect to recording, storing or uploading the video-files. The FTP allowed secure transfer of video files to the central assessor. Following the protocol, there were no need to edit the video-files and the entire process including set-up and transmission took no longer than a few minutes per case.

### **Agreement between local investigators and central assessors**

The cross-tabulation of pair ratings by local investigators and central assessors is represented in **Table 1**. The percentage of total agreement (diagonal cells) was higher using video-based assessments than phone-based assessments for all mRS scores. Globally, total agreement between local and central assessor was obtained in 62.5% of cases (95% CI, 53.2-71.2),  $k_w = 0.77$  (95% CI, 0.72-0.83) using phone calls and in 86.8% of cases (95% CI, 78.8-92.6),  $k_w = 0.92$  (95% CI, 0.88-0.96) using video recordings. In the group of 55 patients that received both central assessments, agreement between both central raters was 67.3% (95% IC 53.3 - 79.3),  $k_w = 0.78$  (95% CI, 0.69-0.88). Compared to local investigator, phone assessor gave a lower score in 22.5% and a higher score in 15% of cases; video-assessor gave a lower or higher score in the same percentage of cases (6.6%), compared to local investigator.

Magnitude and direction of discrepancies in each treatment arm is represented in **Table 2**. Globally, mean difference in mRS scoring between local and central raters (phone and video-based) was comparable regardless of treatment allocation ( $p = 0.3466$ ) but differed in those patients evaluated by video-recordings ( $p = 0.0075$ ). Percentage of direction of discrepancies is represented in **Table 3**.

### **Quality of local mRS interviews as assessed by the central video-based assessor**

Quality of local face-to-face interviews as assessed by central video evaluator were poor in 11/106 cases (10.4%), acceptable in 19/106 cases (17.9%) and reliable in 76/106 (71.7%). The higher the quality of face-to-face interviews, the better the agreement between video-based central assessor and local investigator (36.4%, 84.2% and 92.1% for poor, acceptable and reliable clips respectively). Percentage of poor quality interviews was higher in the first period after implementation of video-recordings compared to the last period (13.3% versus 7.2%).

### **Sensitivity analyses for the primary outcome of REVASCAT trial**

Sensitivity analyses for primary outcome of REVASCAT trial using different adjudication methods of final mRS scores are represented in **Figure 1**. All analyses showed benefit of endovascular treatment but odds ratio slightly differed, being higher when using local evaluations.

### **Association of final infarct volume and mRS scores**

Spearman correlation coefficients between FIV and mRS scores were similar among the three adjudication methods in the group of 55 patients that underwent all evaluations, but differed slightly between treatment arms with local ratings (**Table 4**).

## **Discussion**

We evaluated two different methods of central mRS adjudication within an acute stroke endovascular trial. Both modalities, phone and video-based, were found to be feasible, with no patient compliance related issues or concerns about breach of confidentiality. Central blinded adjudication of mRS provided quality control of mRS interviews at local sites and avoided potential bias regarding blinding. In addition, files containing functional assessment (audio and video clips) were available and stored as back-up copies which



represent an advantage if outcomes are desired to be reanalyzed centrally for the purposes of pooled analyses or for the purposes of conducting analyses using different endpoints.

Reliability of central adjudication was evaluated assessing inter-rater agreement of central assessors with face-to-face evaluations at local sites, the latter being generally considered the standard methodology. Agreement was good for both modalities, but it was higher with video-based adjudication than phone-based adjudications. Agreement between local and central video assessors was also higher compared to CARS study<sup>7</sup>, probably due to the disagreements between the adjudication committee members in CARS study, which occurred in more than 50% of evaluations. Furthermore, in our relatively small sample size, agreement between local and video-based assessors was 100% for mRS scores 0, 3 and 4. Phone assessor gave lower scores (corresponding to lower disability states) than local investigators more frequently than video-based assessor. This may be explained by the fact that central video rater is reviewing the assessment that was performed at the same point in time and in the same environment as the local investigator, receiving exactly the same information from the patient and proxy (if present); this is not the case with phone adjudication which was performed at different time points. In addition, for those patients who are too disabled to provide the information themselves, questions may have been answered by a different proxy. Importantly, the central video assessor is able to observe and evaluate patient global status and some abilities (e.g walking) directly while phone assessor is not. Furthermore behavioral aspects of stroke recovery such as neglect, anosognosia with impact on patient's account of their ability to function can be better assessed by direct visual assessment.

Another advantage of central video-based adjudication is that it provides quality control of face-to-face interviews performed at local sites. Although there were no video-clips unable to be scored, agreement was better when local investigators followed correctly the structured interview, as mandated by the study protocol. In case an investigator was not performing mRS assessment properly as judged by the central assessor, feed-back could be given to local sites to improve training of local investigators for future interviews. Indeed, we observed a reduction in poor quality interviews across the study. Ideally, the expert

feed-back should be given in real time allowing the investigators to re-perform the interview following the central assessor advice. Therefore, the possibility of schedule the out-patient visit with availability of central assessor to perform the evaluation in real time would be of great interest. This could be implemented in future studies thanks to the advent of new technologies as high speed internet, two-way video conferencing and the ubiquitous use of smart phones.

The few observed disagreements between video assessor and local investigators are to be expected given the inherent interrater variability of this adjudication method<sup>12</sup>. However, we must be alert for potential bias due to lack of blinding within endovascular stroke trials in which treatment allocation is open. Therefore, we compared disagreement between local and central raters in both treatment arms. While not significant differences were noted in the whole cohort, in the group of patients evaluated by video, mean differences in mRS scoring varied among treatment arms and were in the direction of benefit of endovascular treatment. Although the reason for these observed discrepancies is unclear, there is a possibility that these differences may in part be explained by lack of blinding by evaluators working at recruiting centers.

The statistical analysis plan in REVASCAT study did not a priori establish which of the three outcome adjudication methods will be the main method used for the primary outcome. Based on the unequivocally blinded nature of the central video adjudication in conjunction with high inter-rater reliability rates between local adjudications and central video adjudications, prior to the first interim analysis the blinded steering committee of REVASCAT decided to consider the central video analysis as primary method used for endpoint adjudication with local evaluation used as default method in case the former was missing. Sensitivity analyses were pre-planned using different adjudication methods for mRS. These analyses reflected the effect of selecting different endpoint adjudication methods on trial results. Had local mRS assessments been chosen as primary outcome adjudication method, the treatment effect of endovascular treatment in REVASCAT would have been higher than reported.

Because all three measurement methods are prone to errors and the true disability status of each patient cannot be ascertained with certainty we sought to validate these three distinct ways of obtaining the mRS against another more objective variable that is known to strongly correlate with neurological recovery, which is 24 hour infarct volume <sup>13</sup>. We found similar and good degree of correlation across all evaluation modalities that did not favor any of the evaluation methods.

The main limitation of the present study is that central adjudication modalities varied along the study (phone in the first period and video in the last), and only 55 patients received both central evaluations. This fact prevented direct comparisons between the two central assessors in the whole cohort and hampered sensitivity analyses of primary outcome of the trial using only one central adjudication method. Also, having the three adjudication methods in the whole cohort of 206 patients would have improved the results statistically and also avoided the problem of unequal numbers in each group. Another limitation is that being only one central assessor in each modality we cannot be certain whether the findings apply to a rating pattern of particular central assessors; however, it represents also a strength point due to the elimination of interobserver variability among different central assessors.

In summary, central adjudication of mRS using video-recordings is feasible and easy to implement in a stroke trial setting, improving quality of assessment of primary outcome and avoiding potential bias. In addition, it confers the advantage of permanent data recording. These results should be considered in the planning of future randomized acute stroke trials, especially in those with open treatment allocation.

## **Sources of funding**

REVASCAT was funded by a local independent Catalan institution (Fundació Ictus Malaltia Vascular, [www.fundacioictus.com/es](http://www.fundacioictus.com/es)) by means of an unrestricted grant from the manufacturer of the device (Covidien). This project has been partially supported by a grant from the Spanish Ministry of Health co-financed by FEDER (Instituto de Salud Carlos III, RETICS-INVICTUS, RD 12/0014/008).

## **Disclosures**

None

## References

1. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int. J. Stroke*. 2009;4:200–205.
2. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified rankin scale: A systematic review. *Stroke*. 2009;40:3393–3395.
3. Wilson JTL, Hareendran A, Hendry A, Potter J, Bone I, Muir KW. Reliability of the modified Rankin Scale across multiple raters: Benefits of a structured interview. *Stroke*. 2005;36:777–781.
4. Quinn TJ, Lees KR, Hardemark HG, Dawson J, Walters MR. Initial experience of a digital training resource for modified Rankin Scale assessment in clinical trials. *Stroke*. 2007;38:2257–2261.
5. Janssen PM, Visser NA, Dorhout Mees SM, Klijn CJM, Algra A, Rinkel GJE. Comparison of telephone and face-to-face assessment of the modified Rankin Scale. *Cerebrovasc. Dis.* 2010;29:137–139.
6. Savio K, Pietra GL Della, Oddone E, Reggiani M, Leone MA. Reliability of the modified Rankin Scale applied by telephone. *Neurol. Int.* 2013;5:6–7.
7. McArthur KS, Johnson PCD, Quinn TJ, Higgins P, Langhorne P, Walters MR, et al. Improving the efficiency of stroke trials: Feasibility and efficacy of group adjudication of functional end points. *Stroke*. 2013;44:3422–3428.
8. Molina C, Chamorro A, Rovira À, de Miquel A, Serena J, Roman L, et al. REVASCAT: A randomized trial of revascularization with SOLITAIRE FR® device vs. best medical therapy in the treatment of acute stroke due to anterior circulation large vessel occlusion presenting within eight-hours of symptom onset. *Int. J. Stroke*. 2015;10: 619–26.
9. Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 Hours after Symptom Onset in Ischemic Stroke. *N Engl J Med*. 2015; 372:2296–306.
10. Saver JL, Filip B, Hamilton S, Yanes A, Craig S, Cho M, et al. Improving the reliability of stroke disability grading in clinical trials and clinical practice: The rankin focused assessment (RFA). *Stroke*. 2010;41:992–995.

11. Kosior JC, Idris S, Dowlatshahi D, Alzawahmah M, Eesa M, Sharma P et al. Quantomo: validation of a computer-assisted methodology for the volumetric analysis of intracerebral haemorrhage. *Int. J. Stroke*. 2011;6:302–305.
12. Quinn TJ, Dawson J, Walters MR, Lees KR. Exploring the Reliability of the Modified Rankin Scale. *Stroke*. 2009;40:762–766.
13. Rangaraju S, Liggins JTP, Aghaebrahim A, Streib C, Sun CH, Gupta R et al. Pittsburgh outcomes after stroke thrombectomy score predicts outcomes after endovascular therapy for anterior circulation large vessel occlusions. *Stroke*. 2014; 45: 2298–2304.



**Figure legend**

**Figure 1.** Sensitivity analyses for primary outcome of the trial (mRS shift) using different adjudication methods.



**Table 1.** Cross-tabulation of pair ratings of mRS by local investigators and central assessors

|                                 |       | Local evaluation (face-to-face interview) |             |             |             |             |             |
|---------------------------------|-------|---|-------------|-------------|-------------|-------------|-------------|
| Central phone (n=120)           |       | 0   | 1           | 2           | 3           | 4           | 5           |
| 0                               | n (%) | 4 ( 80.0%)                                | 7 ( 28.0%)  | 1 ( 4.0%)   |             |             |             |
| 1                               | n (%) | 1 ( 20.0%)                                | 11 ( 44.0%) | 8 ( 32.0%)  |             |             |             |
| 2                               | n (%) |   | 7 ( 28.0%)  | 12 ( 48.0%) | 5 ( 20.0%)  |             |             |
| 3                               | n (%) |   |             | 4 ( 16.0%)  | 19 ( 76.0%) | 2 ( 13.3%)  | 2 ( 8.0%)   |
| 4                               | n (%) |   |             |             | 1 ( 4.0%)   | 8 ( 53.3%)  | 2 ( 8.0%)   |
| 5                               | n (%) |   |             |             |             | 5 ( 33.3%)  | 21 ( 84.0%) |
| Central video-recording (n=106) |       | 0   | 1           | 2           | 3           | 4           | 5           |
| 0                               | n (%) | 7 (100.0%)                                | 4 ( 20.0%)  |             |             |             |             |
| 1                               | n (%) |   | 12 ( 60.0%) |             |             |             |             |
| 2                               | n (%) |   | 3 ( 15.0%)  | 17 ( 85.0%) |             |             |             |
| 3                               | n (%) |   | 1 ( 5.0%)   | 3 ( 15.0%)  | 24 (100.0%) |             |             |
| 4                               | n (%) |   |             |             |             | 11 (100.0%) | 3 ( 12.5%)  |
| 5                               | n (%) |   |             |             |             |             | 21 ( 87.5%) |

**Table 2.** Magnitude of discrepancies among each treatment arm

|                                       |           | <b>Endovascular arm</b> | <b>Medical arm</b> |
|---------------------------------------|-----------|-------------------------|--------------------|
| Difference mRS (central - local)*     | n         | 84                      | 87                 |
|                                       | Mean (SD) | 0.00 (0.51)             | -0.07 (0.53)       |
|                                       | Min, Max  | -1.5, 1.0               | -1.0, 1.0          |
|                                       | p-value   | 0.3466                  |                    |
| Difference mRS (central phone- local) | n         | 57                      | 63                 |
|                                       | Mean (SD) | -0.12 (0.60)            | -0.08 (0.73)       |
|                                       | Min, Max  | -2.0, 1.0               | -2.0, 1.0          |
|                                       | p-value   | 0.7229                  |                    |
| Difference mRS (central video- local) | n         | 55                      | 51                 |
|                                       | Mean (SD) | 0.11 (0.46)             | -0.10 (0.30)       |
|                                       | Min, Max  | -1.0, 2.0               | -1.0, 0.0          |
|                                       | p-value   | 0.0075                  |                    |

Difference in mRS scores between central rater and local rater. \*Central included phone and video-based. In those patients assessed by the two central raters the average score [(phone+video)/2] was used. Mean (SD) represent mean of the difference in mRS scoring (Standard Deviation). Min=minimum difference. Max: maximum difference. P-value is the result of Student T test for the mean differences in scoring between arms

**Table 3.** Direction of discrepancies between local and central raters in each treatment arm

|                                  |       | <b>Endovascular<br/>treatment</b> | <b>Medical arm</b> |
|----------------------------------|-------|-----------------------------------|--------------------|
|                                  | n     | 84                                | 87                 |
| Local scores lower than central  | n (%) | 12 ( 14.3%)                       | 12 ( 13.8%)        |
| Same rating                      | n (%) | 60 ( 71.4%)                       | 56 ( 64.4%)        |
| Local scores higher than central | n (%) | 12 ( 14.3%)                       | 19 ( 21.8%)        |
|                                  | p     | 0.4346                            |                    |

Central scores included phone and video-based. In those patients with both scores the average central score [(phone+video)/2] was used.

**Table 4.** Correlation coefficients (Spearman) between final infarct volume (FIV) and mRS scores

|                      | <b>All patients (n=55)*</b> | <b>Endovascular arm (n=28)</b> | <b>Medical arm (n=27)</b> |
|----------------------|-----------------------------|--------------------------------|---------------------------|
| <b>Local</b>         | 0.458                       | 0.331                          | 0.453                     |
| <b>Central phone</b> | 0.471                       | 0.399                          | 0.426                     |
| <b>Central video</b> | 0.472                       | 0.401                          | 0.437                     |

\*For this analysis only patients with the three evaluations (local, phone and video) were selected to enable comparison of coefficients (same infarct volumes)

